



SFA Portal V1.0

Install Guide

Version 1.2

9 February 2001

SFA Portal Version 1.1.1

9 February 2001

Installing the SFA Portal Software

This document describes the process to install the SFA Portal Software. Modifications to the original document supplied by Exolve/Edge were made during the installation at the VDC in Meriden, CT from 8/23/00 to 8/26/00. The initial install went smoothly on August 23rd, permissions and other “details” were taken care of on subsequent days.

The initial install was performed using /opt/IBMHTTPD/htdocs as the document root directory. This will change when the new development, test, and staging environments are installed and configured.

Note: Three environments will exist concurrently on su35e1 (Sun Solaris Server). These environments will have the following directory structures for three separate installations of the SFA Portal Software:

Development: /opt/dev/IBMHTTPD/htdocs/infospc/SFA
Test: /opt/tst/IBMHTTPD/htdocs/infospc/SFA
Staging: /opt/stg/IBMHTTPD/htdocs/infospc/SFA

The viador.cshrc file must be edited to reflect the installation paths listed above.

1.1 Prerequisites and Configuration

The following items must be done before the SFA Portal software can be successfully installed:

1.1.1 Viador User

A user named “viador” must be created. The Viador user must own the Viador and JRun directories, and will be the user responsible for running JRun. This user will also own the SFA Portal installed code base. This will prevent ownership conflicts from occurring on the source files.

1.1.2 Oracle

Oracle version 8.1.5 must be installed and executing. Fill in the table below:

Oracle Install Directory:	/opt/oracle/product/8.1.5
----------------------------------	---------------------------

1.1.3 IHS (httpd.conf config)

IBM WebSphere version 3.0.2 with IHS version 1.3.6.3 must be installed. Open httpd.conf under /opt/IBMHTTPD/conf, and change or add the following entries:

Alias /admin/ /opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/admin
Alias /SFAPortal/ /opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs
Alias /SFAImages/ /opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/images

In order to display the page you want when entering only the site’s URL in your browser, edit the following line in http.conf by placing the name of the page you wish to display in front of index.html:

Directory Index YourPage.html index.html index.htm (this line will direct the browser to open
YourPage.html first, and index.html second)

WebSphere Install Directory (IHS):	/opt/IBMHTTPD/htdocs
---	----------------------

	Will eventually be: Development - /opt/dev/ IBMHTTPD/htdocs Test - /opt/tst/ IBMHTTPD/htdocs Staging - /opt/stg/ IBMHTTPD/htdocs
--	---

1.1.4 Viador

Viador version 6.1.1 must be installed and executing. The Viador user should own the Viador hierarchy, and all Viador processes should be run as the Viador user. Fill in the table below:

Viador Machine:	Su35e1
Viador Install Directory:	/opt/IBMHTTPD/htdocs/infospc Will eventually be: Development - /opt/dev/ IBMHTTPD/htdocs/infospc Test - /opt/tst/ IBMHTTPD/htdocs/infospc Staging - /opt/stg/ IBMHTTPD/htdocs/infospc
Viador Database Name:	VICDEV
Viador Database User Name:	comm
Viador Database User Password:	*****
Viador Port Number:	6889
Viador infospc directory path:	/opt/IBMHTTPD/htdocs/infospc Will eventually be: Development - /opt/dev/ IBMHTTPD/htdocs/infospc Test - /opt/tst/ IBMHTTPD/htdocs/infospc Staging - /opt/stg/ IBMHTTPD/htdocs/infospc

1.1.5 JRun

The version of JRun that ships with Viador 6.1.1 has a major bug in it, and will not run Java Server Pages correctly. JRun version 2.3 build 157 must be installed instead. It can be installed directly over the Viador version of JRun, which will maintain all necessary configuration information. The Viador user should own and execute all JRun processes and files.

Follow these steps to install JRUN version 2.3 build 157 over an existing version, and to configure it for Schools Portal Application:

- 1) Place patch file (jp23157u.tar.Z) in jrun directory (/opt/disk2/jrun)
- 2) Uncompress jp23157u.tar.Z
- 3) Un-tar jrun tar file in the jrun directory (UNIX:/opt/disk2/jrun>tar xvf jp23157u.tar)
- 4) On an X windows system, type xhost + (allows other machines to open X windows sessions locally)
- 5) As the root user from a C shell, set your Display (UNIX>setenv DISPLAY IP Address of local terminal:0.0)
- 6) Run install.sh as root. Follow the configuration instructions in the Viador installation manual. Except, under the JSP version mapping selection choose "MAP *.jsp to JSP 1.0". Save and exit
- 7) Run startadmin.sh as root (A GUI will come up).
- 8) Highlight jsm, and hit the configure button

- 9) Highlight jse, and hit the service config button
- 10) On the general “tab,” edit the default servlets directory by adding the following path:
/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/classes
- 11) On the file settings tab, unclick the directory browsing allowed button. Save and Close this window
- 12) Choose the General tab, then the java tab. Edit the Java Class by adding these three entries:
/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/classes, /opt/oracle/product/8.1.5/jdbc/lib,
/opt/IBMHTTPD/htdocs/infospc/packages
- 13) On the same tab, add the following to the Shared Library Path: /opt/oracle/product/8.1.5/jdbc/lib

JRun Install Directory:	/opt/disk2/jrun Will eventually be: Development - /opt/dev//disk2/jrun Test - /opt/tst /disk2/jrun Staging - /opt/stg/disk2/jrun
--------------------------------	--

Once the configuration of JRUN is complete, edit the httpd.conf file (/opt/IBMHTTPD/conf) in order to move the newly appended “LoadModule jrun_module136...” line from the end of the file to the “Load Module” portion of the file. This is done to keep the file looking uniform. JRUN sticks this line on the end of the httpd.conf file. The line belongs in the LoadModule portion of the file (to keep things in order).

1.1.6 Autonomy

Autonomy version 2.0 must be installed on the system (there is an Autonomy configuration section later in section 1.4).

Autonomy Install Directory:	/var/autonomy/autonomy_db (on su35e2)
------------------------------------	---------------------------------------

1.1.7 JDK

The Java Development Kit (JD) version 1.1.8 must be installed on the system. This allows for the code to be compiled with the necessary configuration changes for installation.

JDK Install Directory:	/usr/java1.1
-------------------------------	--------------

1.1.8 Servlet API

The Java Servlet API must be installed under the JDK install directory in the lib folder, /usr/java1.1/lib, to allow for compilation of the servlets within the SFA Portal.

1.1.9 System Administration Support

The system administrator must be available to configure, troubleshoot and restart IHS, Oracle, JRun, Autonomy, and Viador if necessary during this install procedure. This install script contains none of these procedures.

1.2 File Installation Procedure

First, the files that make up the SFA Portal must be installed under the Viador hierarchy.

Step	Command
Log in as the Viador user	% su - viador
Create a directory under the Viador infospc Directory called SFA, if it	% cd /opt/IBMHTTPD/htdocs/infospc/dev % mkdir SFA

does not exist.	% cd SFA
Copy the directory structure of the SFA Schools Portal Application from the CD to the SFA directory	% cd /cdrom % cp -R SFA /opt/IBMHTTPD/htdocs/infospc/dev
A new directory, called build, should now be present under the SFA directory. Move into it.	% cd build

1.3 SFA Portal Configuration

The next step is to modify the configuration files within SFA to point to the correct Oracle and Viador instances. This is done by editing the `SfaParameters.java` file and recompiling the Portal.

Step	Command
Move to the directory containing the SFA Portal configuration file	% cd /opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/src/com/edgetech/sfa
Edit the <code>SfaParameters.java</code> file using your favorite text editor.	Vi SfaParameters.java
Set <code>logFileName</code> to the fully specified log file filename, if you do not want to use the default	Use default name
Set <code>httpRoot</code> to the machine name or IP address on which the Portal will be running. Make sure the <code>http://</code> is there.	Use the following IP address: 4.20.14.131
Set the Oracle database parameters to their proper values to access the Viador repository (see the Viador section in Prerequisites, above)	SID: VICDEV User: comm Password: *****
Set the Viador machine name, port number, and infospc directory parameters to their proper values (see the Viador section in Prerequisites, above)	Viador Machine name: 4.20.14.131 Viador Port: 6889 Infospc path: /opt/IBMHTTPD/htdocs/infospc
Save the file and exit the editor	
Go back to the root directory and rebuild the SFA Portal code base	% cd ../../../../ % make -if permissions pose problems, and the above make command does not work, compile the java using the following command from the /opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/src/com/edgetech/sfa directory: UNIX> jc SFAParameters.java

1.4 Autonomy Configuration

The Autonomy configuration files must now be modified to point to the proper web pages for spidering. This is also important to keep synchronization between the SFA Portal search pages and the Autonomy database layout.

The attachment labeled “Summary of Files Modified to Achieve SFA Portal Rendering” is presented to offer an overview of the files which need to be changed. The third attachment is labeled “Highlighted changes to Autonomy Configuration Files.” It offers the entire body of the files, including highlighted changes for the installation of SFA Portals. Make these changes to the files, and place them in the directories specified in the attachment (Highlighted changes to Autonomy Configuration Files).

An additional SFA Schools Portals file must be modified in order to use Autonomy for searching. Proper machine name and filename to direct the web server to the autonomy server must be added to search.jsp. Edit this file (/opt/IBMHTTPD/htdocs/dev/SFA/build/htdocs/search/search.jsp) by changing the path to the old Know Server (search on Edge, or KnowServer), change it to: <http://4.20.14.132/cgi-bin/KnowServer210.exe>

The following files need to be moved to the autonomy box from the SFA Schools Portal web server (they should have gotten to the webserver on install).

The file:

/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/search/query/query.tmpl on su35e1
needs to be moved to:
/var/autonomy/autonomy_db/KnowServer210/wwwtemplates/SFA/query.tmpl on su35e2

The file:

/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/search/results/summary.tmpl on su35e1
needs to be moved to:
/var/autonomy/autonomy_db/KnowServer210/wwwtemplates/SFA/summary.tmpl on su35e2

The file:

/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/search/results/shortresult.tmpl on su35e1
needs to be moved to:
/var/autonomy/autonomy_db/KnowServer210/wwwtemplates/SFA/shortresult.tmpl on su35e2

The file:

/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/search/results/results.tmpl on su35e1
needs to be moved to:
/var/autonomy/autonomy_db/KnowServer210/wwwtemplates/SFA/results.tmpl on su35e2

The file:

/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/search/results/longresult.tmpl on su35e1
needs to be moved to:
/var/autonomy/autonomy_db/KnowServer210/wwwtemplates/SFA/longresult.tmpl on su35e2

The file:

/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/search/index.html on su35e1
needs to be moved to:
/var/autonomy/autonomy_db/KnowServer210/wwwtemplates/SFA/index.html on su35e2

A similar situation results for the images on the autonomy advanced search page. In order to customize the Autonomy advanced search screen, we must take some images from the SFA Schools Portal Web Server and put them on the autonomy server.

The files:

/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/images/corner-right.gif, corner-left.gif,
but_submitl.gif, SFASelect.gif, SFADeselect.gif on su35e1

need to be moved to:

/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/images/corner-right.gif, corner-left.gif,
but_submitl.gif, SFASelect.gif, SFADeselect.gif on su35e2

We must add the following Alias to the httpd.conf file for the IHS installation serving Autonomy (in this case, the IHS installation serving Autonomy was su35e2)

Alias /SFAImages/ /opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/images

1.5 Viador Configuration

From a machine inside the VDC network, enter the following in your Internet Explorer 5.0 browser:

http:// 20.19.67.47/infospc/index.html

(the IP address shown is NAD'ed at the VDC in Meriden, CT... su35e1.ed.gov is valid on the EDLAN)

The Viador login screen will come up. If it does not, either the VIC, IHS, or Oracle are down.

Enter the default Viador username and Password. The Viador application screen will open.

- 1) Click the "USER ADMIN" icon on the right hand portion of the screen
- 2) From the menu bar, *click group -> new group*
- 3) Group Name: SFAUsers
- 4) Description: SP User Group
- 5) Click the tab marked "Sub-Groups," highlight and choose "Everyone"
- 6) Click the tab marked "Roles," highlight and choose "Standard_users"
- 7) Click the tab marked "Document Sources," highlight and choose "Viador Repository"
- 8) Click "OK" to close the window
- 9) From the "USER ADMIN" window, double click on the "Viador" user entry
- 10) Click on the "Parent Groups" tab, and select "SFAUsers"
- 11) Click the "Set Admin Attributes" button
- 12) Choose to "Set as group admin" (you are making the Viador user the admin of all SFAUsers)
- 13) Click "OK"
- 14) Return to main menu
- 15) Double click on the "USER OPTIONS" icon on the right hand portion of the screen
- 16) Edit Preferences of "SFAUsers" by selecting them in the drop down box
- 17) Replace the "Front Page Folder" entry with /dev/SFA/build/htdocs/personal/index.htm
- 18) Click "OK"
- 19) Close Viador, Configuration is complete.

1.6 Installation Complete

At this point, you may begin accessing the SFA Portal. You may wish to back up the files you modified for easier configuration of future versions.

Summary of Files Modified to Achieve SFA Portal Rendering

EdgeKnowledgeServer.cfg

[Configuration]

//This parameter must specify the directory containing the html and template files' root reference.
htmlmdir=/opt/IBMHTTPD/htdocs/en_US/infospc/SFA/build/htdocs/search

//This must be set to 'on' for the suggestion of documents to function correctly.
querysummary=on

[DRE]

//This must specify the address of where the data DRE to use is, most likely the web server.
ipaddress=192.168.99.40
//This represents the query port of the DRE
port=4000
//This is the value of the index port
indexport=4001
//This will specify the number of database known to the DRE (more than this # will be ignored)
databases=13
//Specify the database names to be references by others, including the Fetch and DRE itself
db0 .. db12

EdgeHTTPFetch.cfg

[DEFAULT]

//DRE settings – should point to your data DRE
DREHOST=192.168.99.40
//Index port of data DRE
indexport=4001
//Be sure to keep this value comment out to avoid the Fetch trying to store the original document into the DRE (very bad!)
//IMPORTSTORECONTENT=true
//Represents the number of jobs to be spidered (made to match database names for easy identification)
number=13

[JOB#]

//For each job number use the specific information from the copy of EdgeHTTPFetch.cfg

DRE.INI

[Server]

//The query port for the given DRE
QUERYPORT=4000
//The index port for the given DRE
INDEXPORT=4001

[Databases]

//Specify the total # of databases and their respective names
NUMDBS=13
0= ... 12=

SFA CGI Integration Notes

When adding the IFAP reports to be used as search sources, remember to do the following:

- Update the 'htdocs/search/query.tmpl' file to add the new search sources for querying. Just copy one of the existing examples and paste it in, renaming the database name and number respectively.

Example:

<input type="checkbox"/> BeaconDatabaseReferenceName
--

- Modify 'htdocs/search/search1.htm' to include the new hidden database name and number for form submission in the simple query. `<INPUT TYPE=HIDDEN NAME="dbnum" VALUE="dbname">` This is needed to correctly load the dbase numbers required for the different searches. *See script documentation in search1.htm for more help.*

```

/**
 * This function will need to be modified pending database name/number integration.
 * Based on the selected index of search group ("IFAP & SFA","SFA","IFAP"), it will
 * figure out which hidden fields will be added or removed to correctly submit the query.
 * Each database that is to be used for querying is in the form of :
 *   <input type=hidden name="dbnum" value="dbname">
 * To invalidate the database from being part of the collection, set its value and/or name
 * to null or the empty string ''. The default value is to include only SFA databases
 * that have been set. The names for the new databases will have to be added accordingly.
 * (i.e. when IFAP is chosen all SFA hidden field value have to be nulled out. When SFA
 * is chosen, the IFAP hidden fields have to be nulled out, etc.)
 */
function selectDB() {
    var result = document.forms[0].jumpSelector.selectedIndex
    if (result == 0) { //both IFAP and SFA chosen.
        //Do nothing thus far since all databases should be hard-coded by default.
        (Placeholder)
    } else if (result == 1) { //Only SFA chosen
        //Remove the IFAP references to the CGI
        document.forms[0].db13.value="";
        document.forms[0].db14.value="";
    } else if (result == 2) { //Only IFAP chosen
        //Remove the SFA references into the CGI. Should be a 'for' loop
        //pending database naming/numbering convention integration.
        document.forms[0].db0.value = "";
        //alert(" db0 value is : "+document.forms[0].db0.value);
        document.forms[0].db1.value = "";
        //alert(" db1 value is : "+document.forms[0].db1.value);
        document.forms[0].db2.value = "";
        //alert(" db2 value is : "+document.forms[0].db2.value);
        document.forms[0].db3.value = "";
        //alert(" db3 value is : "+document.forms[0].db3.value);
        document.forms[0].db4.value = "";
        //alert(" db4 value is : "+document.forms[0].db4.value);
        document.forms[0].db5.value = "";
        //alert(" db5 value is : "+document.forms[0].db5.value);
    }
}

```

```

document.forms[0].db6.value = "";
//alert(" db6 value is : "+document.forms[0].db6.value);
document.forms[0].db7.value = "";
//alert(" db7 value is : "+document.forms[0].db7.value);
document.forms[0].db8.value = "";
//alert(" db8 value is : "+document.forms[0].db8.value);
document.forms[0].db9.value = "";
//alert(" db9 value is : "+document.forms[0].db9.value);
document.forms[0].db10.value = "";
//alert(" db10 value is : "+document.forms[0].db10.value);
document.forms[0].db11.value = "";
//alert(" db11 value is : "+document.forms[0].db11.value);
document.forms[0].db12.value = "";
//alert(" db12 value is : "+document.forms[0].db12.value);
}
var goAhead = checkForm();
if (goAhead == true) {
    document.forms[0].submit();
}
}

```

- Let the server know that there are new repositories to extract information from by adding the correct number of databases to the 'KnowledgeServer.cfg' file. *See Adding databases from previous page.*
- Tell the DRE that it has new databases to throw IDX files into. It can't determine from the server which databases it owns, so you have to set it. *See Adding databases in DRE.INI*
- Allow Beacon to add what is needed to complete the integration by adding the necessary lines to the KnowledgeServer.cfg and DRE.INI (or any other files they may need reference to). Pending their API and means of Autonomy interaction, the 2 files listed are the minimum that will be used for correct integration.

Highlighted changes to Autonomy Configuration Files

FILE AND PATH: /var/autonomy/autonomy_db/KnowServer210/KnowServer210.cfg
(originally named EdgeKnowledgeServer.cfg)

Machine: su35e2

[Configuration] (this is where to find the search templates used within the CGI)

htmldir=/opt/IBMHTTPD/htdocs/infospc/dev/SFA/build/htdocs/search
//htmldir=/apps1/Autonomy/Server/wwwtemplates/classic – the old path

autosuggest=on

highlight=sentence

querysummary=on

cgidir=/opt/IBMHTTPD/cgi-bin (must point to the cgi-bin dir)

NumToSuggest=10

HeaderCache=on

SuggestMinThreshold=10

FieldRestrictionOnSuggest=0

highlightStartHTML=

highlightEndHTML=

DateSearch=0

IntelligentSummarySentences=3

[DRE]

ipaddress=4.20.14.131

port=10000

indexport=10001

databases=13

db0=lo_online

db1=current

db2=pellgrantsonline

db3=dlservicer

db4=fafsa

db5=loanconsolidation

db6=topic

db7=sfatech

db8=sfdownload

db9=ifap

db10=eligcert

db11=gapsweb

db12=peps

[HTML]

defaultscreen=main.tmpl

queryinput=query/query.tmpl

database=query/database.tmpl

fields=query/fields.tmpl

date=query/date.tmpl

results=results/results.tmpl

contents=results/contents.tmpl

suggests=results/suggests.tmpl

summary=results/summary.tmpl

noserver=results/noserver.tmpl

noresults=results/noresults.tmpl

nocontents=results/nocontents.tmpl

nosummary=results/nosummary.tmpl

nosuggests=results/nosuggests.tmpl

noneselected=results/noneselected.tmpl

shortresult=results/shortresult.tmpl

longresult=results/longresult.tmpl

suggestresult=results/suggestresult.tmpl

suggestsimilarresults=results/results.tmpl

frameSuggests=results/suggestframe.tmpl

suggestonurlnourl=results/suggestonurlnourl.tmpl

docsubmit=docmanagement/docsubmit.tmpl

docdatabaseselect=docmanagement/databaseselect.tmpl

docsubmitsuccess=docmanagement/submitsuccess.tmpl

doctmplselect=docmanagement/tmplselect.tmpl

[ContentType]

txt=text/plain

html=text/html

htm=text/html

jpg=image/jpeg

jpeg=image/jpeg

gif=image/gif

doc=application/msword

xls=application/vnd.ms-excel

pdf=application/pdf

[Fields]
NumFields=0

[Field0]
Field=title
Mappings=1
Mapping0=title
SearchType=ANYC

[Field1]
Field=NULL
Mappings=1
Mapping0=NULL
SearchType=ANYC

[Field2]
Field=NULL
Mappings=1
Mapping0=NULL
SearchType=ANYC

[DOCMANAGEMENT]
databases=1
db0=AuthoredContent
templates=1
tmpl0=results/contents.tmpl
useDocTemplateField=AuthorTemplate

FILE AND PATH: /var/autonomy/autonomy_db/HTTPFetch210/HTTPFetch210.cfg
(Originally EdgeHTTPFetch.cfg)
Machine: su35e2

[LICENSE]
// license holder details
HOLDER=Autonomy
KEY=616928979192231130733579740051 (may change depending on license)

[SERVICE]
SERVICEPORT=10000
SERVICECONTROLCLIENTS=*
SERVICESTATUSCLIENTS=*

[DEFAULT]
// DRE settings - should point to your document DRE
DREHOST=40.20.14.132 (set to what it needs to be)
INDEXPORT=10001
// DRE Database to use (default) for documents
DATABASE=Archive (should NEVER get here, but things will be put into this db change if none picked)
// Log file to use for logging the detailed request and URL information
LOG FILE=spider.log (keep only during debugging and testing/integration phase – comment out after that)
// Number of sockets to use in parallel. Care if spidering only 1 site
NSOCKETS=32
// Run the spiders every 24 hours

```

SPIDERREPEATSECS=86400
// Run forever
SPIDERCYCLES=-1
// Maximum depth to spider to follow 99 links deep from the front page
DEPTH=99
// Start at 1am
SPIDERSTARTTIME=01:00 (Change this only if have to start service @ diff time)
// Spider for 12 hours MAXIMUM
SITEDURATION=43200
// Only get 100,000 pages maximum
MAXPAGES=100000
// Use SSL Versions 2 & 3 to access https:// prefixed URLs
SECURITYTYPE=SSL_V23
// Strip the text associated with links from a page - it is probably not to do with the page
// but in fact the linked page
IMPORTSTRIPLINKS=ON
// Follow HTTP redirects
FOLLOWREDIRECT=TRUE
// Ensure spiders don't leave the initial URL server
STAYONSITE=TRUE
// Throw out pages smaller than 450 bytes - usually framesets etc.
MINPAGESIZE=450
// Maximum page size 512k
MAXPAGESIZE=512000
// Throw out pages (AFTER following links) if they have more than 50 links - get rid of indices
MAXLINKSPERPAGE=50
// Only wait 60 seconds for data
PAGETIMEOUT=60
// As a news fetch we want no URLs containing the word archive - any case
CANTHAVECHECK=129
CANTHAVECSVS=*archive*
// Some default date formats typically seen
DATEFORMATS=DDMMYYYY,YYMMDD,YYMMD,DDMONTHYYYY,LONGMONTH
DD,MMDDYY,MMDD,DD-SHORTMONTH-YYYY,SHORTMONTH H+:N+:S+ ZZZ YYYY
// Import every 50 pages as they arrive
BATCHPROCESS=INDEX
BATCHSIZE=5
// Extract meta tags into DRE fields - OFF
EXTRACTMETA=OFF
// Be polite to external sites ! you can alter these but beware of the consequences for the //sites you spider
FOLLOWROBOTPROTOCOL=TRUE
// Wait between requests from a given site for 30 seconds to be nice to the remote site!
PAGEDELAY=5

// Import first 3 sentences as summaries
IMPORTSUMMARY=on
// STore textual content in the DRE
// IMPORTSTORECONTENT=true (Do not store content unless you really want to)
// Remove duplicates by reference
INDEXMODE=REFERENCE
// Remove duplicates with a 90% conceptual match or by reference
// INDEXMODE=REFERENCEMATCH90
// Index over socket - allows DRE to be on another machine
INDEXOVERSOCKET=on
// Re member the site structure to make respidering more efficient
storesitestructure=on

```

```
// Now break large documents into conceptual paragraph chunks
IMPORTBREAKING=ON
ImportBreakingMinParagraphWords=300
ImportBreakingMaxParagraphWords=500
ImportBreakingMinDocWords=500
// Don't bother indexing anything less than 10 words!
ImportMinLengthWords=10
// Don't bother indexing anything less than 100 bytes after import
ImportMinLength=100
```

```
[SPIDER] (should be set to the correct default directory out there)
SPIDERPATH=/var/autonomy/autonomy_db/HTTPFetch210
IMPORTPATH=/var/autonomy/autonomy_db/HTTPFetch210
INDEXPATH=/var/autonomy/autonomy_db/HTTPFetch210
```

```
NUMBER=13 (change this number only if adding new databases below)
```

```
0=LO_ONLINE
1=CURRENT
2=PELLGRANTSONLINE
3=DLSEVICER
4=FAFSA
5=LOANCONSOLIDATION
6=TOPIC
7=SFATECH
8=SFADOWNLOAD
9=IFAP
10=ELIGCERT
11=GAPSWEB
12=PEPS
```

```
[LO_ONLINE]
URL=https://lo-online.ed.gov/lo_online/index.shtml
DATABASE=lo_online
SECURITYTYPE=SSL_V23
DATECHECK=0
BEFOREDATE=100000
AFTERDATE=-100000
```

```
[CURRENT]
URL=http://ifap.ed.gov/library/current.htm
DATABASE=current
DEPTH=15
DATECHECK=0
BEFOREDATE=100000
AFTERDATE=-100000
FOLLOWROBOTPROTOCOL=FALSE
```

```
[PELLGRANTSONLINE]
URL=http://www.pellgrantsonline.ed.gov/
DATABASE=pellgrantsonline
DATECHECK=0
BEFOREDATE=100000
AFTERDATE=-100000
```

[DLSERVICER]

URL=<http://www.dlservicer.ed.gov/>
DATABASE=dlservicer
DATECHECK=0
BEFOREDATE=100000
AFTERDATE=-100000

[FAFSA]

URL=<http://www.fafsa.ed.gov/>
DATABASE=fafsa
DATECHECK=0
BEFOREDATE=100000
AFTERDATE=-100000

[LOANCONSOLIDATION]

URL=<http://loanconsolidation.ed.gov/>
DATABASE=loanconsolidation
DATECHECK=0
BEFOREDATE=100000
AFTERDATE=-100000

[TOPIC]

URL=<http://ifap.ed.gov/library/topic.htm>
DATABASE=topic
DEPTH=15
DATECHECK=0
BEFOREDATE=100000
AFTERDATE=-100000

[SFATECH]

URL=<http://ed.gov/offices/OSFAP/sfatech/>
DATABASE=sfatech
DATECHECK=0
BEFOREDATE=100000
AFTERDATE=-100000
FOLLOWROBOTPROTOCOL=FALSE

[SFADOWNLOAD]

URL=<http://www.sfadownload.ed.gov/index.htm>
DATABASE=sfadownload
DATECHECK=0
BEFOREDATE=100000
AFTERDATE=-100000

[IFAP]

URL=<http://ifap.ed.gov/>
DATABASE=ifap
DEPTH=15
DATECHECK=0
BEFOREDATE=100000

AFTERDATE=-100000

[ELIGCERT]

URL=http://www.eligcert.ed.gov/

DATABASE=eligcert

DATECHECK=0

BEFOREDATE=500000

AFTERDATE=-500000

[GAPSWEB]

URL=http://gapsweb.ed.gov/

DATABASE=gapsweb

DATECHECK=0

BEFOREDATE=500000

AFTERDATE=-500000

[PEPS]

URL=http://ed.gov/offices/OSFAP/PEPS/

DATABASE=peps

DATECHECK=0

BEFOREDATE=100000

AFTERDATE=-100000

FOLLOWROBOTPROTOCOL=FALSE

FILE AND PATH: /var/autonomy/autonomy_db/KnowServer210/querydre/DRE.INI

(No name change from original)

Machine: su35e2

[License]

Holder=Autonomy

Key=3169A88B20942368849381Y9010003

[Server]

QUERYCLIENTS=*. *.*.*

INDEXCLIENTS=*. *.*.*

QUERYPORT=10000 (change to the ports used out there before copied over)

INDEXPORT=10001

MAX_DOC_PER_STRCTDAT=20000

ECHO=0

LOGGING=0

QUERYSUMMARY=1

AUTODELETE=1

URLHASH=1

NEWWEIGHT=1

WORDCOUNT=1

SUGGESTTERMS=25

COMBINE=1

HYPHENS=1

HYPHENCHARS=-/

DONTINDEXNUMBERS=1

PROXIMITY=1

TERMSIZE=10
CHARCONV=0
STRIPLANGUAGE=0
NEWCONTENT=0

[Schedule]
COMPACT=100
TIME=23:59
INTERVAL=24:00
EXPIRE=1

[Cache]
MaxTerms=5000
MaxDocs=200
TermThreshold=254
MinFlush=500
MinClear=20

[Databases]
NUMDBS=13 (Modify only if adding new databases below)

0=lo_online
1=current
2=pellgrantsonline
3=dkservicer
4=fafsa
5=loanconsolidation
6=topic
7=sfatech
8=sfdownload
9=ifap
10=eligcert
11=gapsweb
12=peps

[Fields]
NumFields=7
Field0=DREREFERENCE
Field1=INDEX10.DRETITLE
Field2=summary
Field3=CHECKSUM
Field4=DREDOCTYPE
Field5=BIAS
Field6=AuthorTemplate

[Service]
SERVICEPORT=10000
SERVICESTATUSCLIENTS=127.0.0.1
SERVICECONTROLCLIENTS=127.0.0.1